

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 1, 2001, 16:16:41 ; Search time 91.75 Seconds

(Without alignments)
22.175 Million cell updates/sec

Title: US-09-331-631A-3_COPY_186_248

Perfect score: 353
Sequence: 1 KRDPQREYEDRCRRCEQOE.....LINFQGGSGHYEEGEEKOS 63

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 88757 seqs, 32294092 residues

Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	110	31.2	588	VCLB_GOSHI	P09801 gossypium h
2	101.5	28.6	605	VCLB_GOSHI	P09799 gossypium h
3	81	22.9	966	SSNE_YEAST	P14322 saccharomyc
4	78	22.1	573	GLBI_MAIZE	P15590 zea mays (m
5	74	21.0	763	GLIH_CAEEL	P34689 caenorhabdi
6	74	21.0	1090	NIT4_NEUCR	P28349 neurospora
7	73	20.7	386	CYCB_PEA	P13919 pisum sativ
8	73	20.7	919	ANDR_HUMAN	P10275 homo sapien
9	72.5	20.5	185	T2_MOUSE	Q06666 mus musculu
10	71.5	20.3	758	YM38_YEAST	Q07825 saccharomyc
11	71.5	20.3	1898	TRHY_HUMAN	Q07823 homo sapien
12	71	20.1	2442	CBP_HUMAN	Q02832 homo sapien
13	70.5	20.0	550	BLSA_HUMAN	P26805 firlend murti
14	70	19.8	538	GAG_MLYEP	P26805 firlend murti
15	70	19.8	911	ANDR_PANTR	O97775 pan troglod
16	69.5	19.7	975	CDP_CANFA	P39881 canis famli
17	68	19.5	643	GAG_SEV3L	P27400 simian foam
18	68	19.3	47	AGRP_LUECY	P56568 luffa cylin
19	67.5	19.1	164	2SS3_ARATH	P15459 arabidopsis
20	67.5	19.1	1023	GITR_DROME	P33438 drosophila
21	67	19.0	243	AG16_TRYBB	Q26768 lrypanosoma
22	67	19.0	339	TF2D_HUMAN	P20226 homo sapien
23	67	19.0	905	SNF5_YEAST	P18480 saccharomyc
24	66	18.7	429	AP44_MACFA	P23090 duplan murti
25	66	18.7	529	GAG_MLYDU	P29168 murine leuk
26	66	18.7	536	GAG_MLYDE	P29168 murine leuk
27	66	18.7	539	GAG_MLYVO	P26807 firlend murti
28	66	18.7	540	GAG_MLYHO	P21335 homulv murti
29	66	18.7	614	RU17_HUMAN	P08621 homo sapien
30	66	18.7	1154	WCL_NEUCR	O01371 neurospora
31	65.5	18.6	465	HYIN_BRAJA	P19922 bradyrhizob
32	65.5	18.6	656	DNA_STRCO	P27902 streptomyce
33	65.5	18.6	816	ATX1_HUMAN	P54253 homo sapien

34	65.5	18.6	1344	1	IF3A_MOUSE	P23116 mus musculu
35	65.5	18.6	1382	1	IF3A_HUMAN	O14152 homo sapien
36	65	18.4	157	1	HMA1_SCHGR	P29556 schistocerc
37	65	18.4	571	1	CYCA_PEA	P13915 pisum sativ
38	65	18.4	700	1	BIB_DROME	P23645 drosophila
39	65	18.4	1265	1	CYAS_CANFA	P30803 canis famli
40	65	18.4	2476	1	ZAN_PIG	O28983 sus scrofa
41	64.5	18.3	170	1	2SS2_ARATH	P15458 arabidopsis
42	64.5	18.3	807	1	L100_ADE05	P24933 human adeno
43	64.5	18.3	1290	1	DDX8_CAEEL	O09530 caenorhabdi
44	64.5	18.3	1349	1	TRHY_SHEEP	P22793 ovis aries
45	64	18.1	255	1	LP61_EIMTE	P15714 elimeria ten

ALIGNMENTS

RESULT	ID	VCLB_GOSHI	STANDARD	PRT	588 AA.
AC	P09801				
DT	01-MAR-1989 (Rel. 10, Created)				
DT	01-MAR-1989 (Rel. 10, Last sequence update)				
DT	15-JUL-1999 (Rel. 38, Last annotation update)				
DE	VICILIN C72 PRECURSOR (ALPHA-GLOBULIN B).				
OS	Gossypium hirsutum (Upland cotton).				
OC	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RA	Chlan C.A., Pyle J.B., Legocki A.B., Dure L. III;				
RT	"Developmental biochemistry of cottonseed embryogenesis and germination. XVIII. cDNA and amino acid sequences of the members of the storage protein families."				
RL	Plant Mol. Biol. 7:475-489(1986).				
CC	-1- FUNCTION: SEED STORAGE PROTEIN.				
CC	-1- SUBCELLULAR LOCATION: CYTOPLASMIC MEMBRANE-BOUND VACUOLAR PROTEIN BODIES.				
CC	-1- SIMILARITY: NO OTHER 7S SEED STORAGE PROTEINS (PHASEOLIN, VICILIN, CONVICILIN, CONGLICININ, ETC.).				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@sib-sib.ch).				
CC	-----				
DR	EMBL: M16891; AAA33071.1; -				
DR	PIR: A30838; FWCNAB.				
DR	HSSP: P50477; ICAX.				
DR	INTERPRO: IPR001113; -				
DR	PFAM: PF00546; Seedstore_7s; 1.				
FT	Seed storage protein; Signal.				
FT	SIGNAL 1 25				
FT	CHAIN 26 588				
FT	SEQUENCE 588 AA: 69729 MW: 636699829AB8ADEB CRC64;				
FT	VICILIN C72.				
QY	3 DPQOREYEDRCRRCEQDEPRQYQCQRRC-----REDQ-----ROHGRCGDLINQ 48				
Db	82 DPQRR-YEEQOCRCROOEERQRPQCQRCFKRPEQRCQSQSRQFOECQGHCHQDQ-RPE 139				
QY	49 RGS-----GRY-----EEGEEKOS 63				
Db	140 RKQQVRECRERYQENPWRREEREEAEFEETEGDEQOS 178				

Query Match 31.2%; Score 110; DB 1; Length 588;
Best Local Similarity 33.3%; Pred. No. 0.00026;
Matches 33; Conservative 12; Mismatches 14; Indels 40; Gaps 6;

```

RESULT 2
VCLA_GOSHI STANDARD: PRT: 605 AA.
AC P09799:
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE VICILIN GC72-A PRECURSOR (ALPHA-GLOBULIN A).
OS Gossypium hirsutum (Upland cotton).
OC Eukaryota: Viridiplantae: Embryophyta: Tracheophyta: Spermatophyta:
OC Magnoliophyta: eudicotyledons: core eudicots: Rosidae: eurosids II:
OC Malvales: Malvaceae: Gossypium.
(1)
RN SEQUENCE FROM N.A.
RA Chlan C.A., Borroto K., Kamalay J.A., Dure L. III;
RT "Developmental biochemistry of cottonseed embryogenesis and
RT germination. XIX. Sequences and genomic organization of the alpha
RT globulin (vicilin) genes of cottonseed."
RL Plant Mol. Biol. 9:533-546(1987).
CC RT
CC -1- FUNCTION: SEED STORAGE PROTEIN.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC MEMBRANE-BOUND VACUOLAR PROTEIN
CC BOIES.
CC -1- SIMILARITY: TO OTHER 7S SEED STORAGE PROTEINS (PHASEOLIN, VICILIN,
CC CONVICTILIN, CONGLICININ, ETC.).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: M19378: AAA3069.1; -
CC PIR: S06398: S06398.
CC DR HSSP: P50477: ICAX.
CC DR INTERPRO: IPR001113; -
CC DR PFAM: PF00546: Seedstore_7s; 1.
CC DR Seed storage protein; Signal.
CC FT SIGNAL 1 23
CC FT CHAIN 24 605 VICILIN GC72-A.
CC FT SEQUENCE 605 AA: 71049 MW: C9DB9371C976553B CRC64:
SO
Query Match 28.8%; Score 101.5; DB 1; Length 605;
Best Local Similarity 33.3%; Pred. No. 0.002; 7; Indels 21; Gaps 2;
Matches 20; Conservative 12; Mismatches 7; Indels 21; Gaps 2;
QY 3 DPOQREYEDCRRCDEQPRQYOCORCRGROHGRGDLINPORGSGRYEGCEKRO 62
DB 79 DPOQR-YDPCROHCODEERRLRPHCEQSCREO-----YKQKQKQO 117
RESULT 3
SSN6_YEAST STANDARD: PRT: 966 AA.
ID SSN6_YEAST
AC P14922:
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last annotation update)
DE GLUCOSE REPRESSION MEDIATOR PROTEIN.
GN SSN6 OR CYC8 OR YBR112C OR YBR0908.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota: Fungi: Ascomycota: Saccharomycetes: Saccharomycetales:
OC Saccharomycetaceae; Saccharomyces.
(1)
RN SEQUENCE FROM N.A.
RA MEDLINE=89211964; PubMed=2854095;
RA Trumbly R.J.;
RT "Cloning and characterization of the Cyc8 gene mediating glucose
RT repression in yeast."
RL Gene 73:97-111(1988).

```

```

RN [2]
RX SEQUENCE FROM N.A.
RX MEDLINE=88065502; PubMed=3316983;
RA Schultz J., Carlson M.;
RT "Molecular analysis of SSN6, a gene functionally related to the SNF1
RT protein kinase of Saccharomyces cerevisiae."
RL Mol. Cell. Biol. 7:3637-3645(1987).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=92327848; PubMed=1626431;
RA Mannhaupt G., Stucka R., Ehmlé S., Vetter I., Feldmann H.;
RT "Molecular analysis of yeast chromosome II between CWD1 and LYS2: the
RT excision repair gene RAD10 located in this region belongs to a novel
RT group of double-finger proteins."
RL Yeast 8:397-408(1992).
RN [4]
RP TPR REPEATS.
RX MEDLINE=90124639; PubMed=2404612;
RX Sikorski R.S., Boguski M.S., Goebel M., Hieter P.A.;
RT "A repeating amino acid motif in CDC23 defines a family of proteins
RT and a new relationship among genes required for mitosis and RNA
RT synthesis."
RL Cell 60:307-317(1990).
CC -1- FUNCTION: IT IS INVOLVED IN REPRESSION BY A1-ALPHA2 AND ALPHA2 AND
CC IN OTHER SYSTEMS AS A GENERAL REPRESSOR OF TRANSCRIPTION. THIS
CC PROTEIN HAS NO OBVIOUS DNA-BINDING DOMAINS. IT MIGHT NOT INTERACT
CC DIRECTLY WITH DNA BUT WITH DNA-BOUND PROTEINS.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- SIMILARITY: CONTAINS 10 TPR DOMAINS.
CC -1- SIMILARITY: TO YEAST GAL1 AND CCR4.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: M23440: AAA34545.1; -
CC DR EMBL: M17826: AAA35103.1; -
CC DR EMBL: X66247: CAA46973.1; -
CC DR EMBL: X78993: CAA55615.1; -
CC DR EMBL: Z35981: CAA85069.1; -
CC DR PIR: S25365; S25365.
CC DR SGD: S0000316: CYC8.
CC DR INTERPRO: IPR001440; -.
CC DR PFAM: PF00515: TPR; 7.
CC KW Transcription regulation; Repressor; Repeat; TPR domain;
CC NUCLEAR protein.
CC FT DOMAIN 15 30 POLY-GLN.
CC FT REPEAT 46 79 TPR 1.
CC FT REPEAT 80 113 TPR 2.
CC FT REPEAT 114 147 TPR 3.
CC FT REPEAT 150 183 TPR 4.
CC FT REPEAT 187 220 TPR 5.
CC FT REPEAT 224 257 TPR 6.
CC FT REPEAT 258 291 TPR 7.
CC FT REPEAT 296 329 TPR 8.
CC FT REPEAT 330 363 TPR 9.
CC FT REPEAT 364 398 TPR 10.
CC FT DOMAIN 493 556 30 x 2 AA TANDEM REPEATS OF Q-A.
CC FT DOMAIN 557 587 POLY-GLN.
CC FT CONFLICT 547 547 K -> Q (IN REF. 3).
CC FT SEQUENCE 966 AA: 107202 MW: 84B509CF3208C5C0 CRC64:
SO
Query Match 22.9%; Score 81; DB 1; Length 966;
Best Local Similarity 32.0%; Pred. No. 0.35;
Matches 16; Conservative 16; Mismatches 18; Indels 0; Gaps 0;
QY 1 KRDPQREYEDCRRCDEQPRQYOCORCRGROHGRGDLINPORG 50

```


DB 159 QREDEPER 167

RESULT 8
ANDR_HUMAN STANDARD: PRT; 919 AA.

AC P10275;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE ANDROGEN RECEPTOR (DIHYDROTESTOSTERONE RECEPTOR).
GN AR OR NR3C4 OR DHTR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=69112208; PubMed=3216866;
RA Lubahn D.B., Joseph D.R., Sar M., Tan J., Higgs H.N., Larson R.E., French F.S., Wilson E.M.;
RT "The human androgen receptor: complementary deoxyribonucleic acid cloning, sequence analysis and gene expression in prostate.";
RL Mol. Endocrinol. 2:1265-1275(1988).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=90083302; PubMed=2594783;
RA Lubahn D.B., Brown T.R., Simental J.A., Higgs H.N., Migeon C.J., Wilson E.M., French F.S.;
RT "Sequence of the intron/exon junctions of the coding region of the human androgen receptor gene and identification of a point mutation in a family with complete androgen insensitivity.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:9534-9538(1989).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=90258935; PubMed=2342476;
RA Govindan M.V.;
RT "Specific region in hormone binding domain is essential for hormone binding and trans-activation by human androgen receptor.";
RL Mol. Endocrinol. 4:417-427(1990).
RN [4]
RP SEQUENCE FROM N.A.
RC TISSUE=PROSTATE;
RX MEDLINE=89017168; PubMed=3174628;
RA Chang C., Kokontis J., Liao S.;
RT "Structural analysis of complementary DNA and amino acid sequences of human and rat androgen receptors.";
RL Proc. Natl. Acad. Sci. U.S.A. 85:7211-7215(1988).
RN [5]
RP SEQUENCE FROM N.A.
RC TISSUE=PROSTATE;
RX MEDLINE=89098909; PubMed=2911578;
RA Tilley W.D., Marcelli M., Wilson J.D., McPhaul M.J.;
RT "Characterization and expression of a cDNA encoding the human androgen receptor.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:327-331(1989).
RN [6]
RP SEQUENCE FROM N.A.
RC TISSUE=PROSTATE;
RX MEDLINE=91155943; PubMed=2293020;
RA Marcelli M., Tilley W.D., Wilson C.M., Griffin J.E., Wilson J.D., McPhaul M.J.;
RT "Definition of the human androgen receptor gene structure permits the identification of mutations that cause androgen resistance: premature termination of the receptor protein at amino acid residue 588 causes complete androgen resistance.";
RL Mol. Endocrinol. 4:1105-1116(1990).
RN [7]
RP SEQUENCE OF 189-919 FROM N.A.
RX MEDLINE=88178111; PubMed=3353726;
RA Chang C., Kokontis J., Liao S.;
RT "Molecular cloning of human and rat complementary DNA encoding androgen receptors.";
RL Science 240:324-326(1988).

RN [8]
RP SEQUENCE OF 468-919 FROM N.A.
RX MEDLINE=88240407; PubMed=3377788;
RA Trapman J., Klaassen P.C., Kuiper G.G.J.M., van der Korput J.A.G.M., Faber P.M., van Rooij H.C.J., Geurts van Kessel A., Voorhorst M.M., Mulder E., Brinkmann A.O.;
RT "Cloning, structure and expression of a cDNA encoding the human androgen receptor.";
RL Biochem. Biophys. Res. Commun. 153:241-248(1988).
RN [9]
RP POLYMORPHISM OF POLY-GLN REGION.
RX MEDLINE=92220629; PubMed=1561105;
RA Sledzeds H.F., Oostra B.A., Brinkmann A.O., Trapman J.;
RT "Trinucleotide repeat polymorphism in the androgen receptor gene (AR).";
RL Nucleic Acids Res. 20:1427-1427(1992).
RN [10]
RP POLYMORPHISM OF POLY-GLY REGION.
RC TISSUE=BLOOD;
RA Lu J., Daniels M.;
RT Submitted (FEB-1995) to the EMBL/Genbank/DBJ databases.
RN [11]
RP VARIANTS SMA IN POLY-GLN REGION.
RX MEDLINE=91287825; PubMed=2062380;
RA La Spada A.R., Wilson E.M., Lubahn D.B., Harding A.E., Fatschbeck K.H.;
RT "Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy.";
RL Nature 352:77-79(1991).
RN [12]
RP REVIEW ON VARIANTS.
RX MEDLINE=95023089; PubMed=7937057;
RA Patterson M.N., Hughes L.A., Gottlieb B., Pinsky L.;
RT "The androgen receptor gene mutations database.";
RL Nucleic Acids Res. 22:3560-3562(1994).
RN [13]
RP REVIEW ON VARIANTS.
RX MEDLINE=97169385; PubMed=9016528;
RA Gottlieb B., Trifiro M., Lumbroso R., Vasilou D.M., Pinsky L.;
RT "The androgen receptor gene mutations database.";
RL Nucleic Acids Res. 25:158-162(1997).
RN [14]
RP VARIANT LNCAP ALA-877.
RX MEDLINE=91083633; PubMed=2260966;
RA Veldscholte J., Ris-Stalpers C., Kuiper G.G., Jenster G., Berrevoets C., Claassen E., van Rooij H.C.J., Trapman J., Brinkmann A.O., Mulder E.;
RT "A mutation in the ligand binding domain of the androgen receptor of human LNCap cells affects steroid binding characteristics and response to anti-androgens.";
RL Biochem. Biophys. Res. Commun. 173:534-540(1990).
RN [15]
RP VARIANT CAIS MET-866.
RX MEDLINE=91186983; PubMed=2082179;
RA Brown T.R., Lubahn D.B., Wilson E.M., French F.S., Migeon C.J., Corfen J.L.;
RT "Functional characterization of naturally occurring mutant androgen receptors from subjects with complete androgen insensitivity.";
RL Mol. Endocrinol. 4:1759-1772(1990).
RN [16]
RP VARIANT CYS-774.
RX MEDLINE=91310758; PubMed=1856263;
RA Marcelli M., Tilley W.D., Zoppi S., Griffin J.E., Wilson J.D., McPhaul M.J.;
RT "Androgen resistance associated with a mutation of the androgen receptor at amino acid 772 (Arg->Cys) results from a combination of decreased messenger ribonucleic acid levels and impairment of receptor function.";
RL J. Clin. Endocrinol. Metab. 73:318-325(1991).
RN [17]
RP VARIANT CAIS ASN-695 AND HIS-695, AND SEQUENCE OF 629-723 FROM N.A.
RX MEDLINE=92131007; PubMed=1775137;
RA Ris-Stalpers C., Trifiro M.A., Kuiper G.G., Jenster G., Romalo G., Sai T., van Rooij H.C., Kaufman M., Rosenfield R.L., Liao S.;

```

RT      "Substitution of aspartic acid-686 by histidine or asparagine in the
RT      human androgen receptor leads to a functionally inactive protein with
RT      altered hormone-binding characteristics."
RT      Mol. Endocrinol. 5:1562-1569(1991).
RN      181
RP      VARIANTS CAIS AND PAIS.
RX      MEDLINE=93338440; PubMed=1307250;
RA      Balch J.A., Williams D.M., Davies H.R., Brown B.D., Evans B.A.J.,
RA      Hughes I.A., Patterson M.N.;
RT      "Androgen receptor gene mutations identified by SSCP in fourteen
RT      subjects with androgen insensitivity syndrome.";
RL      Hum. Mol. Genet. 1:497-503(1992).
RN      113
RP      VARIANT CAIS VAL-787.
RX      MEDLINE=92235226; PubMed=1569163;
RA      Nakao R., Haji M., Yanase T., Ogo A., Takayanagi R., Katsube T.,
RA      Fukumori Y., Nawata H.;
RT      "A single amino acid substitution (Met-786-->Val) in the steroid-
RT      binding domain of human androgen receptor leads to complete androgen
RT      insensitivity syndrome.";
RL      J. Clin. Endocrinol. Metab. 74:1152-1157(1992).
RN      120
RP      VARIANT LNCAP ALA-877.
RX      MEDLINE=9222955; PubMed=1562539;
RA      Veldscholte J., Berrevoets C.A., Ris-Stalpers C., Kuiper G.G.,
RA      Jenster G., Trapman J., Brinkmann A.O., Mulder E.;
RT      "The androgen receptor in LNCap cells contains a mutation in the
RT      ligand binding domain which affects steroid binding characteristics
RT      and response to antiandrogens.";
RL      J. Steroid Biochem. Mol. Biol. 41:665-669(1992).
RN      121
RP      VARIANT MET-730.
RX      MEDLINE=92335289; PubMed=1631125;
RA      Newmark J.R., Hardy D.O., Tonb D.C., Carter B.S., Epstein J.L.,
RA      Isaacs W.B., Brown T.R., Barlack E.R.;
RT      "Androgen receptor gene mutations in human prostate cancer.";
RL      Proc. Natl. Acad. Sci. U.S.A. 89:6319-6323(1992).
RN      122
RP      VARIANT CAIS VAL-754.
RX      MEDLINE=93372806; PubMed=8103398;
RA      Lobaccaro J.-M., Lombroso S., Ktari R., Dumas R., Sultan C.;
RT      "An exonic point mutation creates a Maelii site in the androgen
RT      receptor gene of a family with complete androgen insensitivity
RT      syndrome.";
RL      Hum. Mol. Genet. 2:1041-1043(1993).
RN      123
RP      VARIANT CAIS ARG-807.
RX      MEDLINE=94108430; PubMed=8281140;
RA      Adeyemo O., Kallio P.J., Palvimo J.J., Kontula K., Jaenke O.A.;
RT      "A single-base substitution in exon 6 of the androgen receptor gene
RT      causing complete androgen insensitivity: the mutated receptor fails
RT      to transactivate but binds to DNA in vitro.";
RL      Hum. Mol. Genet. 2:1809-1812(1993).
RN      124
RP      VARIANT PAIS VAL-743.
RX      MEDLINE=93315568; PubMed=8325932;
RA      Nakao R., Yanase T., Sakai Y., Haji M., Nawata H.;
RT      "A single amino acid substitution (Gly743 --> Val) in the steroid-
RT      binding domain of the human androgen receptor leads to Reifenstein

```

```

DT      01-FEB-1995 (Rel. 31, Created)
DT      01-FEB-1995 (Rel. 31, Last sequence update)
DT      01-NOV-1997 (Rel. 35, Last annotation update)
DE      OCTAPEPTIDE-REPEAT PROTEIN T2.
CN      SRST OR T2.
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
RN      11
RP      SEQUENCE FROM N.A.
RC      STRAIN=BA1B/C; TISSUE=MACROPHAGE;
RX      MEDLINE=93092084; PubMed=1458435;
RA      di Carlo M., Montana G., Romancino D.P., Monteleone D.;
RT      "A mouse repeat sequence conserved in eukaryotic genomes.";
RL      J. Submicrosc. Cytol. Pathol. 24:467-472(1992).
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@sib.ch).
CC      -----
DR      EMBL; X67863; CAA48048.1; .
DR      MGD; MGI:107677; Srst.
KW      Repeat.
FT      DOMAIN
SQ      SEQUENCE 185 AA; 22805 MW; 8995BEC4EC383971 CMC64;

Query Match 20.5%; Score 72.5; DB 1; Length 185;
Best Local Similarity 29.0%; Pred. No. 0.57;
Matches 18; Conservative 15; Mismatches 24; Indels 5; Gaps 2;

Oy      1 KRDPQREYEDCRRCEDQEPHQYQCCRCRCRQROHGRGDLINPQGGSGRVEEKEE 60
Db      96 EREAERQGREAREAREQREORE--RQGREAREQGREAE--RQGREAREQREGER 150
Oy      61 KQ 62
Db      151 QR 152

RESULT 10
ID      YM38 YEAST STANDARD; PRT; 758 AA.
DC      003825.
DT      01-NOV-1997 (Rel. 35, Created)
DT      01-NOV-1997 (Rel. 35, Last sequence update)
DT      01-NOV-1997 (Rel. 35, Last annotation update)
DE      HYPOTHEICAL 85.0 KDA PROTEIN IN HLJ1-SWP2 INTERGENIC REGION.
CN      YMR164C OR YMR520.13C.
OS      Saccharomyces cerevisiae (Baker's yeast).
OC      Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC      Saccharomycetaceae; Saccharomyces.
RN      11
RP      SEQUENCE FROM N.A.
RC      STRAIN=SS288C / AB972;
RA      Hunt S., Bowman S., Barrell B.G., Rajandream M.A.;
RL      Submitted (May-1995) to the EMBL/Genbank/DBJ databases.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@sib.ch).
CC      -----
DR      EMBL; Z49705; CAA89800.1; .
DR      SGD; S0004774; YMR164C.
KW      Hypothetical protein.

```

[illegible]

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L09190; AAA65582.1; -
DR PIR: A45973; A45973.
DR HSSP: P02633; IBOC.
DR MIM: 190370; -
DR INTERPRO: IPR001751; -
DR INTERPRO: IPR02048; -
DR PFAM: PF01023; S_100; 1.
DR PFAM: PF00036; ehand; 1.
DR PROSITE: PS00018; EF_HAND; 1.
DR PROSITE: PS00303; S100_CABP; 1.
KW Repeat; Calcium-binding.
FT DOMAIN 1 91
FT CA_BIND 22 33
FT CA_BIND 62 73
FT DOMAIN 314 390
FT REPEAT 314 326
FT REPEAT 327 339
FT REPEAT 340 351
FT REPEAT 352 364
FT REPEAT 365 377
FT REPEAT 378 390
FT DOMAIN 391 444
FT REPEAT 391 396
FT REPEAT 397 402
FT REPEAT 403 408
FT REPEAT 409 414
FT REPEAT 415 420
FT REPEAT 421 426
FT REPEAT 427 432
FT REPEAT 433 438
FT REPEAT 439 444
FT DOMAIN 444 702
FT DOMAIN 923 1162
FT REPEAT 923 952
FT REPEAT 953 982
FT REPEAT 1013 1012
FT REPEAT 1013 1042
FT REPEAT 1043 1072
FT REPEAT 1073 1102
FT REPEAT 1103 1132
FT REPEAT 1133 1162
FT DOMAIN 1250 1849
FT REPEAT 1250 1752
FT CONFLICT 1794 1801
FT CONFLICT 1857 1857
FT CONFLICT 1880 1880
SQ SEQUENCE 1898 AA; 247219 MW; A74B5947FB62E31D CRC64;

Query Match 20.3%, Score 71.5; DB 1; Length 1898;
Best Local Similarity 37.3%, Pred. No. 5.9;
Matches 22; Conservative 10; Mismatches 24; Indels 3; Gaps 2;

OY 1 KRDPQREYECRRRCEDQEPKQYQCCRRKCEQ--RQHGGGDLINPQSGSGRYEE 57
DB 1723 ERDKRFRFEQLQGRREQLRSQ-ESDRKFRFEQLRQERREQLRPQQRDKYRWEE 1780

RESULT 12
CBP_HUMAN STANDARD: PRT: 2442 AA.
AC Q92793; Q16376; O00147;
JT 15-JUL-1998 (Rel. 36, CT)

```


FT DOMAIN 344 365 LEUCINE-21PPER.
 FT CARBOHYD 85 85 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 138 138 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 550 AA; 64120 MW; 0298938A5FB80F7 CRC64;

Query Match 20.0%; Score 70.5; DB 1; Length 550;
 Best Local Similarity 32.8%; Pred. No. 2.4;
 Matches 20; Conservative 8; Mismatches 16; Indels 17; Gaps 3;

QY 4 PQOREYEDCR-----RCEQOEPRQYQOCRCRCEQORHGRGDLINPQRG 51
 DB 470 PADRVRLCEHHAAAPRGAPRCQGEPPGPP-RGRRSOKRERERGG---PMGG 524

QY 52 S 52
 DB 525 S 525

RESULT 14
 GAG_MLVFP STANDARD; PRT; 538 AA.
 AC P26805;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 01-AUG-1992 (Rel. 23, Last annotation update)
 DE GAG POLYPROTEIN (CORE POLYPROTEIN) [CONTAINS: CORE PROTEINS P15, P12, P30 AND P10].
 GN GAG.
 OS Friend murine leukemia virus (isolate PVC-211) (F-MuLV).
 OC Viruses; Retroviral viruses; Retroviridae; Mammalian type C retroviruses.
 RN (1)
 RX MEDLINE=92319660; PubMed=1620621;
 RA Remington M.P., Hoffman P.M., Ruscelli S.K., Masuda M.;
 RT "Complete nucleotide sequence of a neuropathogenic variant of Friend murine leukemia virus PVC-211."
 RL Nucleic Acids Res. 20:3249-3249(1992).
 CC -1- PTM: SPECIFIC ENZYMOLOGICAL CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
 CC -1- MISCELLANEOUS: THIS PROTEIN IS SYNTHESIZED AS A GAG-POLYPROTEIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M93134; AAA46476.1; -
 CC PIR: S35474; S35474.
 CC INTERPRO: IPR000840; -
 CC INTERPRO: IPR001878; -
 CC INTERPRO: IPR002079; -
 CC INTERPRO: IPR003036; -
 CC PFAM: PF02093; Gag_P30; 1.
 CC PFAM: PF01141; gag_P12; 1.
 CC PFAM: PF00098; zf-CMC; 1.
 CC Coat protein; Core protein; Polyprotein; Nucleoprotein; Myristate; Phosphorylation.
 KW CHAIN 2 131 MATRIX PROTEIN P15.
 FT CHAIN 132 215 RNA BINDING PHOSPHOPROTEIN P12.
 FT CHAIN 216 478 CAPSID PROTEIN P30.
 FT CHAIN 479 538 NUCLEOCAPSID PROTEIN P10.
 FT LIPID 2 2 MYRISTATE (BY SIMILARITY).
 SQ SEQUENCE 538 AA; 61033 MW; 2FF9F97DC79DEBE CRC64;

Query Match 19.8%; Score 70; DB 1; Length 538;
 Best Local Similarity 31.3%; Pred. No. 2.7;
 Matches 21; Conservative 13; Mismatches 13; Indels 20; Gaps 3;

QY 1 KRDPQOREYEDCRRCRCEQOEPRQYQOCRCRCEQORHGRG 41
 DB 433 KRETPERE-ERVRLTEEREERRAEDEREKRRRREMSKILLATVSGRDRG 491

QY 42 GDILNQ 48 -
 DB 492 GERRRQ 498

RESULT 15
 ANDR_PANTR STANDARD; PRT; 911 AA.
 AC 097775;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE ANDROGEN RECEPTOR (DIHYDROTESTOSTERONE RECEPTOR).
 GN AR OR NR3C4.
 OS Pan troglodytes (Chimpanzee).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Pan.
 RN (1)
 RX MEDLINE=98404153; PubMed=9732460;
 RA Choong C.S., Kempainen J.A., Wilson E.M.;
 RT "Evolution of the primate androgen receptor: a structural basis for disease."
 RL J. Mol. Evol. 47:334-342(1998).
 CC -1- FUNCTION: THE STEROID HORMONES AND THEIR RECEPTORS ARE INVOLVED IN THE REGULATION OF EUKARYOTIC GENE EXPRESSION AND AFFECT CELLULAR PROLIFERATION AND DIFFERENTIATION IN TARGET TISSUES.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.
 CC -1- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN, A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.
 CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
 CC NR3 SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U94177; AAC73048.1; -
 CC HSP: P06536; IRGD.
 CC INTERPRO: IPR000536; -
 CC INTERPRO: IPR001103; -
 CC INTERPRO: IPR001628; -
 CC PFAM: PF00104; hormone_rec; 1.
 CC PFAM: PF00105; zf-C4; 1.
 CC PRINTS: PR00047; STEROIDRECEPTOR.
 CC PROSITE: PS00031; NUCLEAR_RECEPTOR; 1.
 CC Receptor; Transcription regulation; DNA-binding; Nuclear protein; Zinc-finger; Steroid-binding.
 KW DOMAIN 1 549 MODULATING (BY SIMILARITY).
 FT DOMAIN 551 616 C4-TYPE ZINC FINGERS (TWO).
 FT ZN_FING 551 571 C4-TYPE.
 FT ZN_FING 587 611 C4-TYPE.
 FT DOMAIN 682 911 LIGAND-BINDING.
 FT DOMAIN 57 78 POLY-GLN.
 FT DOMAIN 84 88 POLY-GLN.
 FT DOMAIN 192 196 POLY-GLN.
 FT DOMAIN 371 380 POLY-PRO.
 FT DOMAIN 395 401 POLY-ALA.
 FT DOMAIN 448 464 POLY-GLY.
 SQ SEQUENCE 911 AA; 98402 MW; 601B9BD4E697DAA4 CRC64;

Query Match 19.8%; Score 70; DB 1; Length 911;

